Jurnal Impresi Indonesia (JII)
e-ISSN: 2810-062x
Vol. 5, No. 10, October 2026, pp 6154-6172
DOI : 10.58344/jii.v5i10.8170

Optimasi Integrasi XGBoost dan Bayesian Optimization untuk Peningkatan Akurasi Prediksi Diabetes

Rahmat Sutandya Yudhanto*, Sri Huning Anwariningsih, Farid Fitriyadi
Universitas Sahid Surakarta, Indonesia
Email : rsutandya.yudhanto@gmail.com*, srihuning@usahidsolo.ac.id
faridfitriyadi@gmail.com

Keywords:
XGBoost, Bayesian Optimization, Optimization, Prediction, Diabetes

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Riviera Publishing

Abstract
The increasing prevalence of diabetes highlights the need for accurate prediction models, while class imbalance limits the identification of individuals with prediabetes. This study evaluated the integration of XGBoost and Bayesian optimization to improve diabetes classification using population health surveillance data. A quantitative experimental design used the Behavioral Risk Factor Surveillance System (BRFSS) dataset for 2023–2025, comprising 782,941 observations, 21 features, and three diabetes status categories. A temporal split assigned 2023 data to training, 2024 data to validation, and 2025 data to testing. Preprocessing included deduplication and imputation fitted exclusively on training data. Optuna optimized hyperparameters across 40 trials by maximizing macro-F1. Class weighting addressed imbalance, with SMOTE examined through sensitivity analysis. Test accuracy increased from 68.91% to 73.21%, while macro-F1 increased from 49.88% to 50.85%. However, prediabetes recall declined from 31% to 24%, indicating that improvements primarily reflected better identification of individuals without diabetes. During validation, class weighting achieved a macro-F1 of 0.515, compared with 0.382 when combined with SMOTE. These findings demonstrate improved aggregate performance without resolving minority class recognition. The integration improves overall prediction, but targeted strategies remain necessary before clinical application. Future research should investigate cost-sensitive learning, staged classification, probability calibration, and stability across subsequent periods.

Kata Kunci:
XGBoost, Bayesian Optimization, Optimasi, Prediksi, Diabetes

Abstrak
Peningkatan prevalensi diabetes mendorong kebutuhan model prediksi yang akurat, sementara ketidakseimbangan kelas membatasi pengenalan kelompok pradiabetes. Penelitian ini bertujuan mengevaluasi integrasi XGBoost dan Bayesian Optimization untuk meningkatkan kinerja klasifikasi diabetes pada data survei kesehatan. Penelitian menggunakan pendekatan kuantitatif eksperimental dengan dataset Behavioral Risk Factor Surveillance System (BRFSS) periode 2023–2025, mencakup 782.941 observasi awal, 21 fitur, dan tiga kategori status diabetes. Data dibagi berdasarkan waktu, yaitu tahun 2023 untuk pelatihan, 2024 untuk validasi, dan 2025 untuk pengujian. Prapemrosesan meliputi deduplikasi dan imputasi berbasis data latih. Optimasi hiperparameter dilakukan menggunakan Optuna melalui 40 percobaan dengan memaksimalkan macro-F1. Ketidakseimbangan kelas ditangani melalui pembobotan kelas, disertai uji sensitivitas menggunakan SMOTE. Hasil pengujian menunjukkan peningkatan akurasi dari 68,91% menjadi 73,21% dan macro-F1 dari 49,88% menjadi 50,85%. Namun, recall pradiabetes

menurun dari 31% menjadi 24%, sehingga peningkatan terutama berasal dari pengenalan kelas tidak diabetes. Pada validasi, pembobotan kelas menghasilkan macro-F1 sebesar 0,515, sedangkan kombinasinya dengan SMOTE hanya mencapai 0,382. Temuan ini menunjukkan bahwa optimasi meningkatkan kinerja agregat, tetapi belum mengatasi kelemahan pengenalan kelas minoritas. Disimpulkan bahwa integrasi XGBoost dan Bayesian Optimization memperbaiki prediksi, tetapi pengembangan strategi khusus pradiabetes diperlukan sebelum pemanfaatan klinis. Penelitian lanjutan perlu mengeksplorasi pembelajaran sensitif biaya, klasifikasi bertahap, kalibrasi probabilitas, dan stabilitas temporal.

PENDAHULUAN

Diabetes telah menjadi tantangan utama kesehatan global saat ini, dengan prevalensi yang meningkat pesat setiap tahunnya. Berdasarkan data yang dipublikasikan oleh World Health Organization (WHO) pada tahun 2024 lebih dari 800 juta orang dewasa di seluruh dunia hidup dengan diabetes, sehingga menjadikannya sebagai prioritas kesehatan global untuk di atasi karena dampaknya yang luas (Organization, n.d.). Indonesia menempati peringkat ke-5 dunia untuk jumlah penderita diabetes dan menghadapi tantangan diabetes yang serius dengan prevalensi 11,3% yang mencakup lebih dari 20,4 juta orang dewasa. Penyakit diabetes disebabkan oleh resistensi insulin, penurunan produksi insulin, atau keduanya sehingga kadar gula darah dalam tubuh terjadi peningkatan (Federation, 2025). Tingginya angka kasus diabates yang tidak terdiagnosis juga menjadi tantangan tersendiri, sehingga deteksi dini sangat diperlukan untuk pencegahan serta meminimalisir risiko diabetes (Harefa & Waruwu, 2025).

Faktor risiko diabetes bersifat multidimensi, meliputi indikator klinis dan perilaku seperti hipertensi, obesitas yang tercermin pada BMI, status kesehatan umum, usia, jenis kelamin, riwayat penyakit jantung, aktivitas fisik, dan kebiasaan merokok (Mudjanarko et al., 2023). Keseluruhan indikator tersebut terdokumentasi konsisten dalam *Behavioral Risk Factor Surveillance System* (BRFSS) dan banyak dipakai sebagai parameter prediksi pada studi-studi pemodelan risiko diabetes, sehingga menjadikan BRFSS sumber data yang relevan untuk deteksi dini berbasis analitik. Di sisi lain, beban ekonomi akibat diabetes terus meningkat tajam, pengeluaran kesehatan global terkait diabetes diperkirakan melampaui USD 1 triliun pada tahun 2024 sehingga memperkuat urgensi pendekatan prediktif yang mampu mengidentifikasi populasi berisiko tinggi secara lebih akurat untuk intervensi dini. Namun, pemodelan pada data surveilans populasi kerap menghadapi masalah ketidakseimbangan kelas, tanpa penanganan yang memadai, model dapat mencapai akurasi tinggi tetapi gagal mengenali kasus diabetes (Alamsyah et al., 2025; Li et al., 2024).

Perkembangan rekam medis elektronik dan surveilans kesehatan populasi menciptakan himpunan data berskala besar yang membuka peluang analitik prediktif untuk deteksi dini diabetes. Hal ini membuka peluang akan kebutuhan pendekatan komputasi yang efisien dan berbasis data besar untuk mendukung pengambilan keputusan kesehatan publik. Oleh karenanya pendekatan gabungan berbasis *machine learning* dan teknik optimasi diperlukan untuk meningkatkan sensitivitas dan kegunaan klinis model prediksi (Belle et al., 2015; Cahyana et al., 2025; Khurshid et al., 2025; Kliment et al., 2018; Verma et al., 2025).

Algoritma *machine learning* khususnya XGBoost banyak diadopsi untuk prediksi risiko diabetes karena kemampuannya menangani data kompleks dalam jumlah besar. Meski akurasi agregat sering tinggi, kinerja pada kelas minoritas kerap melemah akibat ketidakseimbangan kelas, studi berbasis BRFSS menunjukkan XGBoost dapat mencapai akurasi >86% namun dengan recall kasus diabetes hanya 17%, sehingga berisiko bias ke kelas mayoritas

(Brahmandjati et al., 2024; Demirci & Yıldırım, 2025; Munshi, 2025; Shindy et al., 2025). Optimasi hiperparameter yang efisien dilaporkan dapat melampaui *random search* dalam menemukan konfigurasi XGBoost yang lebih kuat secara sistematis. Temuan-temuan tersebut menjadi dasar penelitian ini untuk mengkaji sejauh mana integrasi XGBoost dan Bayesian Optimization efektif meningkatkan akurasi dan sensitivitas prediksi diabetes pada dataset BRFSS multi-tahun (2023-2025) dengan 782.941 data dan 22 atribut, termasuk mengidentifikasi sejauh mana penanganan ketidakseimbangan kelas mampu memperbaiki pengenalan kelas minoritas. Pemrosesan data, pemodelan baseline, penyetelan hiperparameter berbasis Bayesian, hingga evaluasi model pada penelitian ini dilakukan pada platform komputasi berbasis cloud Google Colaboratory.

Beberapa penelitian terdahulu memiliki keterkaitan dengan penelitian ini dan menjadi dasar penting dalam penelitian yang diusulkan. Salah satu penelitian menilai efektivitas XGBoost yang disetel secara sistematis untuk klasifikasi diabetes dengan membandingkan beberapa varian XGBoost yang ditingkatkan melalui grid search dan Bayesian optimization terhadap sejumlah algoritma pembanding. Hasilnya menunjukkan akurasi yang sangat tinggi dengan keunggulan XGBoost pada akurasi, F1, dan MCC, sehingga menegaskan bahwa penyetelan hiperparameter mampu memberikan lonjakan kinerja di atas model dasar (Barman et al., 2025; Fitriyadi & Indriyati, 2025; Rizal et al., 2025). Penelitian tersebut memiliki persamaan dengan penelitian ini pada penggunaan machine learning untuk prediksi diabetes, namun berbeda pada ruang lingkup yang masih berfokus pada dataset tunggal sehingga kajian generalisasi yang lebih luas masih diperlukan.

Sejalan dengan upaya optimasi tersebut, penelitian lain melakukan penggabungan optimasi XGBoost berbasis algoritma genetika dengan pendekatan stacking ensemble, dilengkapi prosedur pra-pemrosesan berupa penanganan ketidakseimbangan kelas menggunakan oversampling seperti SMOTE, ADASYN, dan SMOTE-ENN, serta pemberian penjelasan model melalui SHAP pada data BRFSS 2015 rilis Kaggle. Hasilnya menunjukkan peningkatan kinerja yang konsisten setelah optimasi serta membuka peluang riset lanjutan pada seleksi fitur dan strategi optimasi yang lebih efisien. Persamaannya dengan penelitian ini terletak pada tujuan peningkatan akurasi prediksi diabetes melalui optimasi, sedangkan perbedaannya terletak pada pendekatan gabungan multi-model serta penggunaan dataset rilis yang lebih lama.

Selain optimasi model, penanganan ketidakseimbangan kelas juga menjadi perhatian pada penelitian yang menekankan pentingnya penanganan ketidakseimbangan kelas pada pemodelan diabetes dengan menerapkan SMOTE bersama XGBoost dan penyetelan hiperparameter melalui RandomizedSearchCV. Penelitian tersebut mencapai akurasi 74,6% dengan recall 83%, sehingga memberikan penekanan pada pengurangan false negative sebagai prioritas klinis. Persamaannya terletak pada fokus mitigasi ketidakseimbangan kelas dan kebutuhan penyetelan hiperparameter, sedangkan perbedaannya adalah strategi penyetelan yang bersifat acak serta orientasi metrik utama pada recall.

Pada sisi lain, penelitian yang membandingkan strategi optimasi random, grid, dan Bayesian pada model decision tree, random forest, serta XGBoost menyajikan himpunan hiperparameter terbaik, termasuk konfigurasi XGBoost yang diperoleh melalui Bayesian optimization, disertai analisis interpretabilitas model. Persamaannya dengan penelitian ini adalah penekanan pada efisiensi Bayesian optimization dalam menavigasi ruang hiperparameter, sementara perbedaannya terletak pada sifat penelitian yang komparatif lintas algoritma dan tidak menggunakan dataset multi-tahun.

Adapun penelitian lain yang berfokus pada efektivitas pipeline optimasi menunjukkan bahwa kombinasi tuner seperti Optuna dengan algoritma alternatif seperti CatBoost, serta teknik resampling hibrida seperti SMOTE-ENN, dapat mendorong peningkatan akurasi pada klasifikasi diabetes. Persamaannya dengan penelitian ini adalah perhatian pada tuning dan pengendalian ketidakseimbangan kelas sebagai kunci perbaikan kinerja, sedangkan perbedaannya adalah fokus utama yang tidak menggunakan XGBoost sebagai model dasar.

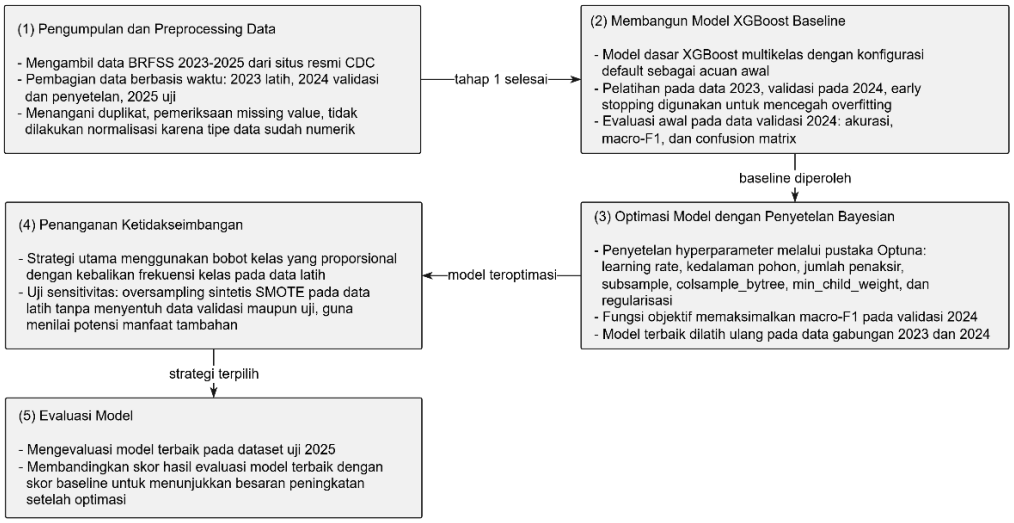
Secara keseluruhan, penelitian-penelitian terdahulu tersebut menegaskan bahwa XGBoost memiliki potensi unggul untuk prediksi diabetes dan kinerjanya sangat dipengaruhi oleh penyetelan hiperparameter serta penanganan ketidakseimbangan kelas. Namun demikian, banyak studi masih terbatas pada dataset dalam satu tahun sehingga aspek generalisasi multi-tahun belum teruji kuat. Penelitian ini menutup kekosongan tersebut dengan mengintegrasikan XGBoost dan Bayesian optimization menggunakan dataset BRFSS dari tahun 2023 hingga 2025 agar diperoleh model yang lebih akurat dan stabil pada skala data populasi.

Berdasarkan penelitian terdahulu, banyak studi memakai dataset tunggal satu tahun atau rilis lama seperti BRFSS 2015 dan 2021 dari Kaggle, sehingga generalisasi lintas waktu belum teruji secara memadai. Walaupun teknik optimasi terbukti meningkatkan XGBoost, penanganan ketidakseimbangan kelas belum dilaporkan secara konsisten, sehingga sensitivitas terhadap kelas diabetes kerap menurun pada skenario populasi besar seperti surveilans kesehatan multi-tahun. Berangkat dari celah-celah yang teridentifikasi pada penelitian terdahulu, penelitian ini menempatkan kontribusinya sebagai berikut: (1) menggunakan data multi-tahun yaitu BRFSS 2023-2025 untuk generalisasi yang lebih luas, (2) membangun baseline XGBoost yang realistis untuk data besar serta imbalanced, (3) menerapkan *Bayesian Optimization* dengan penyetelan hiperparameter yang terstruktur, bukan ansambel berat sehingga dapat menunjukkan selisih kinerja baseline dan teroptimasi secara jelas. (4) mengkaji efektivitas strategi penanganan ketidakseimbangan kelas, termasuk pembobotan kelas dan uji sensitivitas SMOTE terhadap kasus diabetes, (5) seluruh *pipeline* berjalan di Google Colaboratory agar replikasi mudah dan sumber daya terkontrol.

Dengan demikian, penelitian ini menghadirkan kajian yang terukur mengenai integrasi XGBoost dan *Bayesian Optimization* serta penanganan ketidakseimbangan kelas pada data survei kesehatan multi-tahun, sehingga dapat memberikan gambaran objektif mengenai sejauh mana pendekatan tersebut mampu meningkatkan performa prediksi diabetes secara efisien pada komputasi, sekaligus mengidentifikasi batasan yang perlu diperbaiki sebelum model dipertimbangkan sebagai bagian dari sistem pendukung keputusan pada layanan kesehatan, termasuk fasilitas dengan keterbatasan sumber daya.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental untuk membangun dan mengevaluasi model klasifikasi diabetes berbasis XGBoost pada data BRFSS 2023–2025. Alur lengkap penelitian disajikan pada Gambar 1.



Gambar 1. Diagram alur penelitian

Pengumpulan dan *Preprocessing* Data

Tahap awal penelitian ini dimulai dengan pengambilan data BRFSS tahun 2023–2025 dari situs resmi *Centers for Disease Control and Prevention* (CDC). Dataset yang dikumpulkan berjumlah 782.941 data secara keseluruhan, terdiri atas 21 fitur indikator risiko kesehatan dan 1 label target multikelas, dengan rincian jumlah data per tahun yaitu 243.651 data pada tahun 2023, 264.291 data pada tahun 2024, dan 274.999 data pada tahun 2025. Data yang telah di ambil disimpan dalam format .csv, kemudian digabungkan dan diproses melalui serangkaian tahap *preprocessing* untuk menjamin kualitas dan konsistensi data sebelum digunakan pada tahap pemodelan.

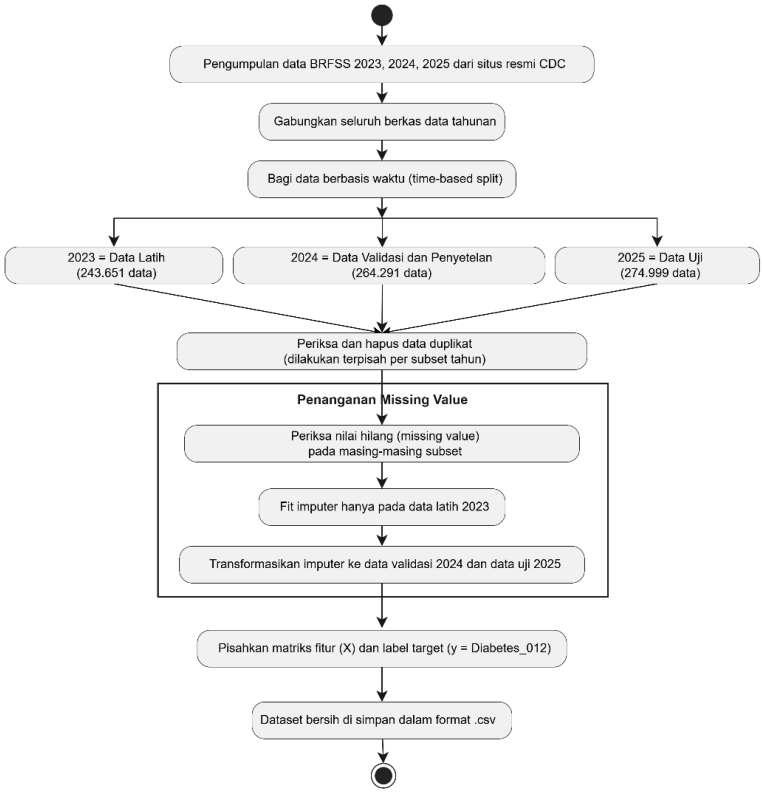
Tahap *preprocessing* diawali dengan pembagian data berbasis waktu (*time-based split*), yaitu data tahun 2023 digunakan sebagai data latih, tahun 2024 sebagai data validasi dan penyetelan hiperparameter, serta tahun 2025 sebagai data uji. Skema ini dipilih untuk menjaga prinsip *anti-leakage*, sehingga tidak ada informasi dari periode yang lebih baru yang memengaruhi proses pelatihan maupun penyetelan model, sehingga evaluasi akhir benar-benar dapat mencerminkan kemampuan generalisasi model pada periode data berikutnya.

Selanjutnya, dilakukan pemeriksaan dan penanganan data duplikat serta nilai hilang (*missing value*) pada masing-masing subset. Penghapusan data duplikat dilakukan secara terpisah pada tiap subset 2023, 2024, dan 2025 untuk mencegah bias akibat pengulangan observasi yang sama, normalisasi tidak dilakukan karena seluruh fitur berbentuk numerik. Penanganan *missing value* dilakukan dengan tetap menjaga prinsip *anti-leakage*, yaitu dengan cara *imputer* di *fit* hanya menggunakan data latih tahun 2023, kemudian digunakan untuk mentransformasikan data validasi 2024 dan data uji 2025, sehingga tidak ada informasi dari data validasi maupun uji yang ikut memengaruhi proses imputasi.

Tahap terakhir pada *preprocessing* adalah pemisahan matriks fitur (X) dan label target (y). Label target yang digunakan adalah *Diabetes_012*, yang merepresentasikan tiga kategori kondisi, yaitu 0 untuk tidak diabetes, 1 untuk prediabetes, dan 2 untuk diabetes. Pemilihan *Diabetes_012* sebagai target didasarkan pada kebutuhan penelitian untuk membangun model klasifikasi multikelas yang mampu membedakan tidak hanya status diabetes secara biner, melainkan juga mengenali kondisi prediabetes sebagai kelas peralihan, sehingga model yang

dihasilkan memiliki nilai deteksi dini yang lebih relevan secara klinis dibandingkan pendekatan klasifikasi biner.

Dari tahap pengumpulan dan *preprocessing* data ini diperoleh dataset yang bersih dan terstruktur, siap digunakan untuk proses latih, validasi, dan uji. Diagram alur tahapan pengumpulan dan *preprocessing* data serta dokumentasi ringkas statistik fitur serta pemetaan indeks baris setiap tahun disajikan pada Gambar 2 dan 3.



Gambar 2. Diagram alur tahapan pengumpulan dan *preprocessing* data

Subset (Tahun)	Rentang Indeks Baris	Jumlah Baris	Fitur	Mean	Std	Min	Max
2023	0 – 238284	238285	BMI	28.35	4.89	12.0	51.75
2023	0 – 238284	238285	Age	9.22	3.07	1.0	13.00
2023	0 – 238284	238285	GenHlth	2.70	1.14	1.0	5.00
2024	0 – 257458	257459	BMI	28.51	4.89	12.0	51.74
2024	0 – 257458	257459	Age	9.22	3.06	1.0	13.00
2024	0 – 257458	257459	GenHlth	2.70	1.14	1.0	5.00
2025	0 – 269466	269467	BMI	28.62	4.89	12.0	53.40
2025	0 – 269466	269467	Age	9.21	3.07	1.0	13.00
2025	0 – 269466	269467	GenHlth	2.70	1.14	1.0	5.00

Gambar 3. Dokumentasi ringkas statistik fitur serta pemetaan indeks baris setiap tahun

Membangun Model XGBoost (Baseline)

Pada tahap ini dibangun model dasar XGBoost untuk klasifikasi multikelas dengan fungsi objektif *multi:softprob*, sehingga keluaran model berupa probabilitas untuk masing-

masing kelas *Diabetes_012*. Konfigurasi default digunakan sebagai acuan awal kinerja, dengan tujuan menyediakan titik pembanding yang objektif sebelum dilakukan optimasi hiperparameter pada tahap berikutnya. Pelatihan model dilakukan menggunakan data latih tahun 2023, sedangkan data tahun 2024 dimanfaatkan sebagai data evaluasi internal untuk mekanisme *early stopping*, dimana ketika peningkatan performa terhadap data validasi tidak lagi signifikan maka proses pelatihan akan dihentikan secara otomatis, sehingga mencegah model mengalami *overfitting* akibat iterasi yang berlebihan. Untuk menjaga reprodusibilitas eksperimen, seluruh proses pelatihan baseline dijalankan pada lingkungan Google Colaboratory dengan *random seed* yang tetap, sementara waktu pelatihan dan jumlah iterasi terbaik hasil *early stopping* turut dicatat sebagai bagian dari dokumentasi eksekusi model (lihat Tabel 4).

Evaluasi awal terhadap model baseline dilakukan pada set validasi 2024, mencakup metrik akurasi, macro-F1, dan *confusion matrix*, guna memperoleh gambaran menyeluruh mengenai kemampuan model dalam mengenali setiap kelas sebelum dibandingkan dengan hasil pada data uji 2025. Dengan demikian, tahap ini menghasilkan skor baseline yang terukur dan terdokumentasi dengan baik, yang selanjutnya menjadi titik referensi untuk menilai besaran peningkatan kinerja setelah penyetelan hiperparameter berbasis Bayesian dilakukan.

Optimasi Model dengan Penyetelan Bayesian

Setelah baseline diperoleh, tahap berikutnya berfokus pada peningkatan kinerja model melalui penyetelan hiperparameter menggunakan *Bayesian Optimization* dengan bantuan pustaka Optuna. Tujuan utama pada tahap ini bukan sekadar meningkatkan akurasi, melainkan meningkatkan kemampuan model dalam mengenali seluruh kelas secara lebih seimbang, sehingga fungsi objektif optimasi ditetapkan untuk memaksimalkan nilai macro-F1 pada data validasi 2024. Ruang pencarian hiperparameter dirancang mengikuti komponen yang paling berpengaruh terhadap kompleksitas dan regularisasi model, meliputi laju belajar (*learning_rate*), kedalaman pohon (*max_depth*), jumlah penaksir (*n_estimators*), *subsample*, *colsample_bytree*, *min_child_weight*, serta regularisasi L1 (*reg_alpha*) dan L2 (*reg_lambda*). Parameter-parameter tersebut secara khusus mengatur keseimbangan antara kapasitas model, digunakan juga strategi *sampling* fitur untuk mengurangi *overfitting*, serta regularisasi agar model tidak terlalu menyesuaikan diri pada pola *noise* dalam data survei.

Proses optimasi dilakukan dengan skema yang tetap menjaga prinsip *anti-leakage*, yaitu model dilatih menggunakan data latih tahun 2023, kemudian dievaluasi pada data validasi tahun 2024 untuk menghasilkan nilai macro-F1 sebagai parameter kualitas tiap *trial*. Agar proses pelatihan lebih efisien, pada setiap *trial* diterapkan mekanisme *early stopping*, sehingga pelatihan akan berhenti ketika kinerja terhadap validasi tidak lagi signifikan dan juga jumlah iterasi efektif setiap *trial* dapat berbeda-beda sesuai karakteristik kombinasi hiperparameter yang diuji. Setelah kombinasi hiperparameter terbaik diperoleh dari keseluruhan proses pencarian, model kemudian dilatih ulang pada data gabungan 2023 dan 2024 untuk disiapkan menuju tahap pengujian akhir pada data 2025. Dengan pendekatan ini, tahap optimasi menghasilkan konfigurasi hiperparameter terbaik yang menjadi dasar pembangunan model teroptimasi pada tahap selanjutnya.

Penanganan Ketidakseimbangan

Dataset BRFSS yang digunakan pada penelitian ini memiliki karakteristik ketidakseimbangan distribusi kelas pada label *Diabetes_012*, sebuah kondisi yang umum terjadi pada data survei kesehatan populasi. Ketidakseimbangan ini berpotensi membuat model bias terhadap kelas mayoritas, sehingga performa keseluruhan tampak baik dari sisi akurasi,

tetapi kemampuan mengenali kelas minoritas menjadi rendah. Untuk menangani masalah tersebut, strategi utama yang digunakan adalah pembobotan kelas (*class weighting*) melalui *sample_weight*, di mana sampel dari kelas minoritas diberikan bobot yang proporsional dengan kebalikan frekuensi kemunculannya pada data latih, sehingga model terdorong untuk mempelajari pola kelas minoritas secara lebih memadai tanpa perlu mengubah struktur maupun jumlah data asli. Pendekatan ini dipilih sebagai strategi utama karena relatif stabil untuk diterapkan pada data survei berskala besar dan tidak mengubah struktur distribusi fitur.

Sebagai uji sensitivitas untuk memastikan bahwa strategi yang dipilih benar-benar sesuai bagi karakteristik dataset ini, dilakukan perbandingan antara dua strategi, yaitu penggunaan *class weight* saja dan kombinasi *class weight* dengan teknik *oversampling* sintetis SMOTE (*Synthetic Minority Oversampling Technique*). SMOTE diterapkan dengan cara membentuk sampel sintetis baru pada kelas minoritas melalui interpolasi antar-titik data yang berdekatan dalam ruang fitur. Uji ini dilakukan secara terkontrol dengan tetap menjaga prinsip *anti-leakage*, yaitu SMOTE hanya diterapkan pada data latih tahun 2023, sementara penilaian kinerja tetap dilakukan pada data validasi 2024, sehingga data uji 2025 tidak pernah tersentuh dalam proses pemilihan strategi ini untuk menjaga objektivitas evaluasi akhir. Berdasarkan hasil perbandingan tersebut, kemudian dipilih konfigurasi penanganan ketidakseimbangan kelas yang memberikan kinerja paling baik dan paling stabil, yang kemudian akan digunakan pada tahap pembangunan model teroptimasi dan evaluasi akhir.

Evaluasi Model

Tahap akhir penelitian ini adalah mengevaluasi model terbaik hasil optimasi pada dataset uji tahun 2025, yang sepenuhnya independen dari proses pelatihan, validasi, maupun penyetelan hiperparameter pada tahap-tahap sebelumnya. Evaluasi dilakukan menggunakan dua metrik utama, yaitu akurasi dan macro-F1, macro-F1 dipilih sebagai metrik karena dapat memperlakukan setiap kelas secara setara tanpa terpengaruh oleh dominasi kelas mayoritas, sehingga lebih relevan untuk menilai kemampuan model dalam mengenali kelas minoritas seperti pradiabetes. Untuk memperdalam analisis kesalahan prediksi antar kelas, hasil evaluasi turut divisualisasikan menggunakan *confusion matrix* yang dinormalisasi per baris, sehingga proporsi prediksi untuk setiap kelas aktual dapat diinterpretasikan secara lebih jelas dan memudahkan identifikasi pola kesalahan yang paling sering terjadi.

Skor hasil evaluasi model terbaik ini kemudian dibandingkan secara langsung dengan skor baseline pada data uji yang sama, guna menunjukkan besaran peningkatan kinerja yang dihasilkan oleh integrasi Bayesian Optimization dan penanganan ketidakseimbangan kelas.

HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil penelitian yang berkaitan dengan proses pengembangan dan peningkatan performa model prediksi diabetes menggunakan algoritma XGBoost serta teknik optimasi hyperparameter berbasis Bayesian Optimization. Pembahasan disusun secara sistematis mulai dari tahap pengumpulan dan preprocessing data BRFSS 2023–2025, pembangunan model baseline, proses penyetelan hyperparameter untuk memaksimalkan nilai macro-F1 pada data validasi, penanganan ketidakseimbangan kelas melalui *class weight*, hingga evaluasi akhir dan analisis perbandingan performa model pada data uji tahun 2025. Seluruh tahapan dipaparkan secara rinci melalui tabel dan visualisasi guna menunjukkan efektivitas pendekatan yang diusulkan dalam meningkatkan akurasi dan konsistensi kinerja model dibandingkan baseline.

Pengumpulan dan *Preprocessing* Data

Dataset yang digunakan pada penelitian ini adalah data Behavioral Risk Factor Surveillance System (BRFSS) tahun 2023–2025, yang diambil melalui situs resmi CDC. Data terdiri atas 21 fitur faktor risiko kesehatan dan 1 label multikelas Diabetes_012 yang merepresentasikan tiga kategori kondisi 0 = tidak diabetes, 1 = prediabetes, 2 = diabetes. Untuk evaluasi yang adil serta menghindari kebocoran informasi (*anti-leakage*), pembagian data dilakukan berbasis waktu dengan urutan tahun, yaitu 2023 sebagai data latih, 2024 sebagai data validasi serta penyetelan, dan 2025 sebagai data uji. Pembagian ini memastikan seluruh pengembangan dilakukan tanpa mengganggu data uji, sehingga evaluasi akhir dapat mencerminkan kemampuan generalisasi model pada data periode berikutnya.

Setelah data dipisahkan menjadi tiga subset tersebut, dilakukan pemeriksaan kualitas data untuk mengidentifikasi nilai hilang (*missing values*) dan duplikat pada masing-masing subset. Ringkasan kondisi awal dataset sebelum pembersihan, khususnya total nilai hilang dan duplikasi ditampilkan pada Tabel 1.

Tabel 1. Ringkasan kualitas data awal per subset pada BRFSS 2023–2025

Subset (Tahun)	Jumlah Baris	Nilai Hilang (Baris)	Duplikat (Baris)
2023	243.651	5001	5366
2024	264.291	5401	6832
2025	274.999	5643	5532

Berdasarkan Tabel 1, tingkat nilai hilang pada ketiga subset tergolong konsisten, yaitu berkisar 2,04–2,05% dari total baris tiap tahun, yang mengindikasikan pola *missing data* yang stabil dan tidak menunjukkan anomali antar periode survei. Sebaliknya, tingkat duplikasi menunjukkan variasi yang lebih besar, yaitu 2,20% pada 2023, meningkat menjadi 2,58% pada 2024, kemudian menurun menjadi 2,01% pada 2025, yang mengindikasikan bahwa duplikasi data kemungkinan besar dipengaruhi oleh proses pengumpulan tahunan yang tidak sepenuhnya seragam antar periode survei BRFSS. Karena baris duplikat dapat menyebabkan bias akibat pengulangan observasi yang sama, proses deduplikasi dilakukan secara terpisah pada masing-masing subset 2023, 2024, dan 2025. Hasil jumlah duplikasi yang berhasil dihapus pada tiap subset ditampilkan pada Tabel 2.

Tabel 2. Ringkasan Jumlah data duplikat yang dihapus

Subset (Tahun)	Sebelum (Baris)	Sesudah (Baris)	Deduplikasi (Baris)
2023	243.651	238.285	5366
2024	264.291	257.459	6832
2025	274.999	269.467	5532

Selanjutnya, penanganan nilai hilang dilakukan dengan prinsip anti-leakage, yaitu imputer di-fit hanya menggunakan data latih tahun 2023, kemudian digunakan untuk mentransformasikan data validasi 2024 dan data uji 2025. Setelah imputasi, nilai hilang pada ketiga subset menjadi nol, sehingga seluruh fitur dapat digunakan dalam pemodelan tanpa mengurangi jumlah baris data. Untuk memberikan gambaran ringkas mengenai kondisi dataset setelah preprocessing, profil dataset per tahun disajikan pada Tabel 3.

Tabel 3. Profil dataset setelah preprocessing

Subset (Tahun)	Jumlah Baris	Jumlah Fitur	Fitur Hilang	Label 0 (Baris)	Label 1 (Baris)	Label 2 (Baris)
2023	238285	21	0	200573	4354	33358
2024	257459	21	0	215871	4932	36656
2025	269467	21	0	225890	5416	38161

Profil dataset pada Tabel 3 juga mengungkap tingkat ketidakseimbangan kelas yang cukup ekstrem, di mana rasio antara kelas 0 (tidak diabetes) terhadap kelas 1 (pradiabetes) mencapai sekitar 46:1 pada 2023 dan berangsur menurun menjadi sekitar 42:1 pada 2025, sementara rasio kelas 0 terhadap kelas 2 (diabetes) relatif stabil pada kisaran 6:1. Proporsi kelas 1 menunjukkan tren peningkatan yang konsisten dari 1,83% pada 2023 menjadi 2,01% pada 2025, sejalan dengan proporsi kelas 0 yang sedikit menurun dari 84,17% menjadi 83,81%. Pola ini secara tidak langsung merefleksikan tren epidemiologis peningkatan prevalensi prediabetes pada populasi survei, sekaligus menegaskan urgensi kemampuan model dalam mengenali kelas minoritas secara lebih baik.

Sebagai luaran akhir tahap preprocessing, dataset yang telah bersih disimpan dalam format CSV untuk kemudian digunakan pada tahap pengembangan model integrasi XGBoost dan optimasi *hyperparameter* berbasis bayesian.

Membangun Model XGBoost (Baseline)

Pada tahap ini, dibangun model baseline sebagai acuan pembanding untuk menilai efektivitas optimasi Bayesian pada tahap berikutnya. Model baseline menggunakan algoritma XGBoost untuk klasifikasi multikelas dengan fungsi objektif *multi:softprob*, sehingga keluaran model berupa probabilitas untuk masing-masing kelas Diabetes_012: 0, 1, 2. Pelatihan model dilakukan menggunakan data latih tahun 2023, sedangkan data tahun 2024 dimanfaatkan sebagai data evaluasi internal untuk mekanisme early stopping agar proses pelatihan berhenti ketika peningkatan performa tidak lagi signifikan. Untuk menjaga reprodusibilitas eksperimen, seluruh proses dijalankan pada lingkungan Google Colaboratory dengan random seed yang tetap, serta waktu pelatihan dan jumlah iterasi terbaik dicatat sebagai bagian dari penelitian ini. Rincian konfigurasi eksekusi baseline berupa waktu pelatihan dan jumlah iterasi terbaik hasil early stopping ditampilkan pada Tabel 4.

Tabel 4. Ringkasan konfigurasi eksekusi dan waktu pelatihan model XGBoost baseline.

Model (Baseline)	Waktu Pelatihan (Detik)	Jumlah Iterasi Terbaik
XGBoost	53,389	170

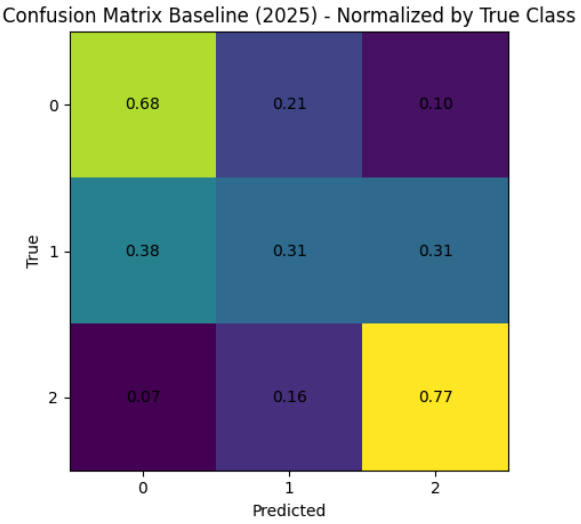
Pada penelitian ini, pelatihan baseline memerlukan waktu sekitar 53,39 detik, dengan jumlah iterasi terbaik 170. Informasi ini penting untuk menunjukkan bahwa baseline dibangun secara terukur dan dapat direplikasi pada lingkungan yang sama. Selanjutnya, performa baseline dievaluasi pada data uji tahun 2025 untuk memperoleh metrik utama penelitian, yaitu akurasi dan macro-F1. Hasil evaluasi baseline pada 2025 disajikan pada Tabel 5.

Tabel 5. Hasil evaluasi model XGBoost baseline pada data uji 2025

Metrik	Precision	Recall	F1 Score	Support
0	0,97	0,68	0,80	225890,0
1	0,03	0,31	0,05	5416,00

2	0,55	0,77	0,64	38161,00
Accuracy	0,69	0,69	0,69	0,69
Macro avg	0,51	0,59	0,50	269467,0
Weighted avg	0,89	0,69	0,76	269467,0

Secara keseluruhan, baseline menghasilkan akurasi 0,69 dan macro-F1 0,50 pada data uji 2025. Berdasarkan hasil tersebut juga terlihat bahwa baseline mencapai F1 Score yang tinggi pada kelas 0 (0,80) dan cukup baik pada kelas 2 (0,64), namun sangat rendah pada kelas 1 (0,05). Pola ini menjadi indikasi awal bahwa kelas 1 (pradiabetes) merupakan kelas yang paling sulit dikenali, yang akan dikaitkan dengan ketidakseimbangan kelas pada tahap selanjutnya. Ketimpangan ini juga tercermin dari kesenjangan antara nilai precision (0,03) dan recall (0,31) pada kelas 1, yang mengindikasikan bahwa mayoritas prediksi tersebut keliru, sehingga menghasilkan *false positive*. Pola kesalahan ini semakin diperjelas melalui confusion matrix (lihat Gambar 4), yang menunjukkan bahwa kesalahan klasifikasi pada kelas pradiabetes tersebar hampir merata ke dua arah, yaitu 38% terklasifikasi sebagai tidak diabetes dan 31% terklasifikasi sebagai diabetes. Temuan ini mengindikasikan bahwa model baseline belum mampu membedakan batas antara ketiga kondisi secara konsisten. Untuk memperjelas pola kesalahan prediksi antar kelas, hasil evaluasi baseline divisualisasikan menggunakan confusion matrix ternormalisasi per baris. Confusion matrix model baseline pada data uji 2025 disajikan pada Gambar 4.



Gambar 4. Confusion matrix model baseline pada data uji 2025

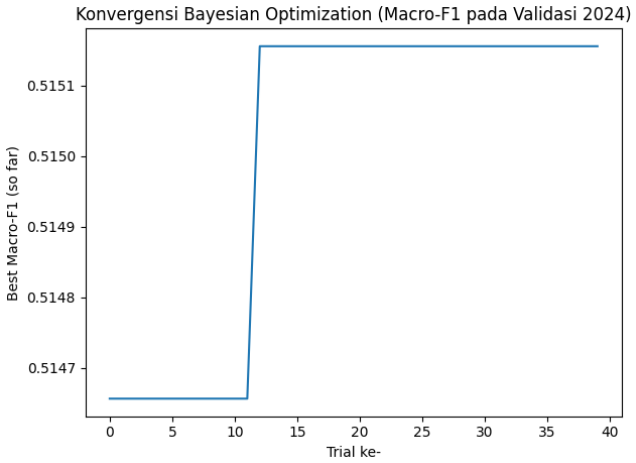
Optimasi Model dengan Penyetelan Bayesian

Setelah baseline diperoleh sebagai titik acuan, tahap berikutnya adalah melakukan optimasi model melalui penyetelan hyperparameter menggunakan *Bayesian Optimization* dengan pustaka *Optuna*. Tujuan utama pada tahap ini bukan sekadar menaikkan akurasi, melainkan meningkatkan kemampuan model dalam mengenali seluruh kelas secara lebih seimbang. Oleh karena itu, fungsi objektif optimasi ditetapkan untuk memaksimalkan nilai macro-F1 pada data validasi tahun 2024.

Proses optimasi dilakukan dengan skema yang tetap menjaga prinsip *anti-leakage*, yaitu model dilatih menggunakan data latih tahun 2023, kemudian dievaluasi pada data validasi

tahun 2024 untuk menghasilkan nilai macro-F1 sebagai parameter kualitas tiap *trial*. Agar proses pelatihan lebih efisien, pada setiap *trial* diterapkan mekanisme early stopping, sehingga pelatihan akan berhenti ketika kinerja pada validasi tidak lagi membaik. Dengan demikian, jumlah iterasi efektif setiap *trial* dapat berbeda-beda sesuai karakteristik kombinasi hyperparameter yang diuji.

Ruang pencarian hyperparameter dirancang mengikuti komponen yang paling berpengaruh terhadap kompleksitas dan regularisasi model. Parameter yang ditelusuri mencakup learning_rate, max_depth, n_estimators, subsample, colsample_bytree, min_child_weight, reg_alpha, dan reg_lambda. Parameter-parameter tersebut mengatur keseimbangan antara kapasitas model, strategi *sampling* fitur untuk mengurangi overfitting, serta regularisasi agar model tidak terlalu menyesuaikan diri terhadap noise. Dengan ruang pencarian ini, *Bayesian Optimization* diarahkan untuk menemukan konfigurasi yang memberikan kinerja validasi paling stabil tanpa meningkatkan risiko overfitting secara berlebihan. Ringkasan proses optimasi ditunjukkan melalui kurva konvergensi pada Gambar 5.



Gambar 5. Kurva konvergensi *Bayesian Optimization* (*Optuna*) berdasarkan macro-F1 pada data validasi 2024

Sedangkan, untuk ringkasan hasil 10 *trial* terbaik dari proses optimasi penyetelan hyperparameter menggunakan *Bayesian Optimization* disajikan pada Tabel 6.

Tabel 6. Ringkasan 10 *trial* terbaik hasil *bayesian optimization*

Trail	Value	State	user_best_iteration	user_train_seconds
12	0,515	COMPLETE	724	280,036
32	0,515	COMPLETE	111	67,248
0	0,515	COMPLETE	421	178,333
22	0,514	COMPLETE	205	104,186
34	0,514	COMPLETE	271	119,594
19	0,514	COMPLETE	76	53,256
21	0,514	COMPLETE	64	48,922
4	0,514	COMPLETE	2200	824,077
31	0,513	COMPLETE	67	49,221
27	0,513	COMPLETE	195	94,802

Tabel 7. Konfigurasi hyperparameter terbaik hasil *bayesian optimization*

Metrik	Nilai
learning_rate	0,005
max_depth	10
n_estimators	4928
subsample	0,857
colsample_bytree	0,704
min_child_weight	0,514
reg_alpha	0,0002
reg_lambda	0,012
best_macro_f1_val_2024	0,515
n_trials	40
tuning_seconds	8468,670

Berdasarkan hasil optimasi (lihat Tabel 7), diperoleh nilai terbaik macro-F1 validasi 2024 sebesar 0,515, dengan konfigurasi hyperparameter terbaik sebagai berikut: (1) learning_rate = 0,005, (2) max_depth = 10, (3) n_estimators = 4928, (4) subsample = 0,857, (5) colsample_bytree = 0,704, (6) min_child_weight = 0,514, (7) reg_alpha = 0,0002, (8) reg_lambda = 0,012.

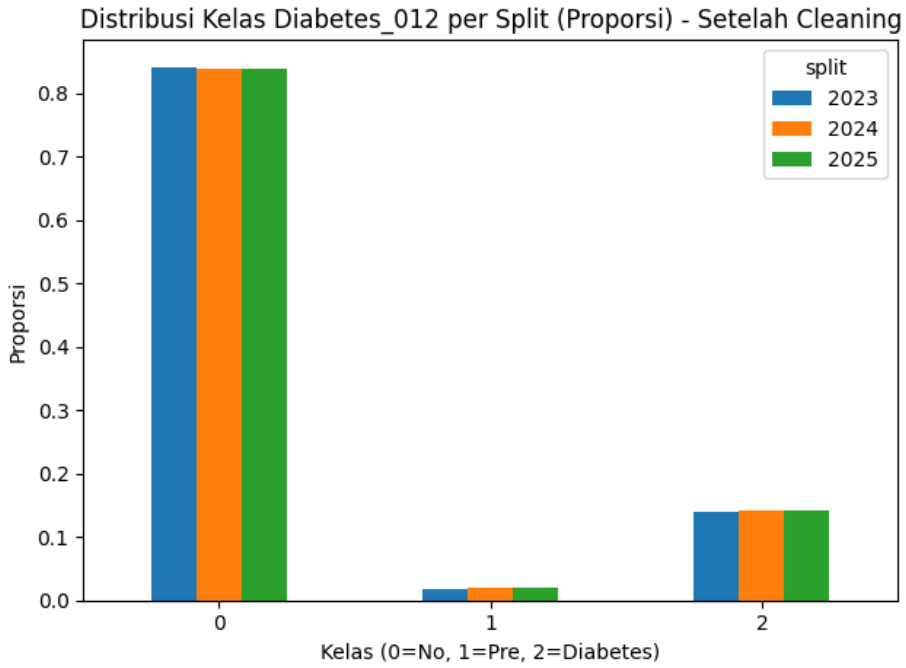
Konfigurasi terbaik cenderung menggunakan *learning rate* yang relatif kecil dan kedalaman pohon yang tinggi, yang mengindikasikan bahwa model memerlukan pembelajaran bertahap namun cukup ekspresif untuk memisahkan pola risiko diabetes pada data survei. Namun, kompleksitas tersebut tetap dikendalikan melalui kombinasi *subsampling* serta regularisasi agar model tidak mudah overfit. Dari sisi komputasi, total waktu tuning tercatat sebesar 8468,670 detik (sekitar 2 jam 21 menit), dengan rata-rata waktu pelatihan sekitar 181,97 detik per trial pada 10 *trial* terbaik dan rata-rata best_iteration 433,4. Parameter terbaik ini kemudian digunakan sebagai dasar untuk membangun model teroptimasi pada tahap berikutnya.

Kurva konvergensi (lihat Gambar 5) menunjukan bahwa peningkatan nilai macro-F1 terjadi secara signifikan pada rentang *trial* awal (0 hingga sekitar *trial* ke-10), sebelum akhirnya melandai dan relatif stabil pada kisaran 0,5151 hingga *trial* ke-40. Pola ini mengindikasikan bahwa algoritma Bayesian Optimization berhasil menemukan konfigurasi optimum secara efisien pada tahap awal pencarian. Hal ini turut didukung (lihat Tabel 6), di mana kesepuluh *trial* terbaik menunjukkan nilai macro-F1 yang berdekatan (0,513–0,515), mengindikasikan keberadaan wilayah hiperparameter yang secara konsisten memberikan kinerja optimum, bukan sekadar satu titik puncak yang rentan terhadap perubahan kecil parameter. *Trial* ke-4 yang memerlukan waktu pelatihan terlama yaitu 824,08 detik dengan 2.200 iterasi, justru menghasilkan macro-F1 yang lebih rendah sebesar 0,5141 dibandingkan *trial* ke-12 yang hanya membutuhkan 280,04 detik dengan 724 iterasi namun mencapai nilai tertinggi sebesar 0,5152. Temuan ini menegaskan bahwa durasi pelatihan yang lebih panjang tidak selalu berkorelasi dengan kinerja yang lebih baik, sekaligus membuktikan efisiensi kombinasi *early stopping* dan pencarian berbasis Bayesian dibandingkan strategi penyetelan *brute-force*.

Penanganan Ketidakseimbangan

Dataset BRFSS yang digunakan pada penelitian ini memiliki karakteristik yaitu ketidakseimbangan distribusi kelas pada label Diabetes_012. Kondisi ini umum terjadi pada data survei kesehatan populasi, di mana pada penelitian ini mayoritas responden berada pada

kelas 0, sementara kelas 1 muncul dalam proporsi jauh lebih kecil, dan kelas 2 berada di antara keduanya. Ketidakseimbangan ini berpotensi membuat model bias pada kelas mayoritas, sehingga performa keseluruhan tampak baik dari sisi akurasi, tetapi kemampuan mengenali kelas minoritas menjadi rendah. Gambaran komposisi label pada masing-masing subset setelah preprocessing divisualisasikan pada Gambar 6.



Gambar 6. Distribusi kelas Diabetes_012 pada masing-masing subset setelah preprocessing.

Untuk menangani masalah tersebut, strategi utama yang digunakan adalah pembobotan kelas (*class weighting*) melalui sample_weight. Pada pendekatan ini, sampel dari kelas minoritas diberikan bobot lebih besar selama proses pelatihan, sehingga model terdorong untuk mempelajari pola kelas minoritas. Pendekatan ini dipilih sebagai strategi utama karena relatif stabil untuk data survei berskala besar dan tidak mengubah struktur distribusi fitur.

Untuk memastikan bahwa strategi yang dipilih adalah yang paling sesuai untuk dataset ini, dilakukan uji sensitivitas dengan membandingkan dua strategi yaitu class_weight dan class_weight dengan SMOTE. Uji ini dilakukan secara terkontrol dengan prinsip *anti-leakage*: (1) SMOTE hanya diterapkan pada data latih 2023, (2) penilaian kinerja dilakukan pada data validasi 2024. Dengan demikian, pemilihan strategi penanganan ketidakseimbangan tidak pernah menyentuh data uji 2025, sehingga tetap menjaga objektivitas evaluasi akhir. Hasil perbandingan kedua strategi pada validasi 2024 disajikan pada Tabel 8.

Tabel 8. Perbandingan strategi penanganan ketidakseimbangan kelas

Strategi	Akurasi	Macro-F1	F1 Kelas 0	F1 Kelas 1	F1 Kelas 2
Class Weight	0,773	0,515	0,864	0,052	0,630
Class Weight + SMOTE	0,392	0,382	0,493	0,039	0,615

Hasil uji sensitivitas (lihat tabel 8) menunjukkan bahwa pendekatan *class weight* memberikan kinerja yang paling baik dan konsisten. Nilai macro-F1 untuk *class_weight* mencapai 0,515, sedangkan strategi *class_weight* dengan SMOTE justru turun drastis menjadi 0,382. Temuan ini mengindikasikan bahwa pada data BRFSS ini, teknik oversampling sintetis seperti SMOTE tidak selalu menguntungkan. SMOTE bekerja dengan cara membentuk sampel baru melalui interpolasi pada ruang fitur, pada dataset yang memuat banyak data diskrit atau biner pada fiturnya, proses ini dapat menghasilkan kombinasi nilai yang kurang representatif dan berpotensi menambah noise, sehingga model kehilangan kemampuan generalisasi. Penurunan performa akibat penerapan SMOTE tergolong drastis, khususnya pada akurasi yang menurun hingga 49,3% secara relatif dari 0,773 menjadi 0,392, jauh melampaui penurunan pada macro-F1 yang tercatat sebesar 25,8% secara relatif. Temuan yang lebih mengkhawatirkan adalah F1 Score pada kelas 1 (prediabetes), kelas yang seharusnya paling diuntungkan oleh teknik *oversampling* justru ikut menurun dari 0,052 menjadi 0,039. Kondisi ini mengindikasikan bahwa pada dataset dengan proporsi kelas minoritas yang sangat kecil sebagaimana ditunjukkan pada Gambar 6 dimana kelas 1 hanya berkisar 1,8–2,0% dari total data setiap tahun serta didominasi oleh fitur-fitur diskrit atau biner, mekanisme interpolasi SMOTE antar-titik data cenderung menghasilkan sampel sintetis yang tidak merepresentasikan pola sesungguhnya, sehingga alih-alih memperkuat kelas minoritas, teknik ini justru menambah *noise* yang merusak kemampuan generalisasi model secara menyeluruh. Oleh karena itu, strategi final yang dipilih adalah *class_weight* tanpa SMOTE. Dengan demikian, penanganan ketidakseimbangan pada penelitian ini ditempatkan sebagai bagian yang tidak terpisahkan. Ketimpangan kelas dibuktikan melalui distribusi label, kemudian ditangani melalui pembobotan kelas sebagai strategi utama, dan diverifikasi melalui uji sensitivitas.

Evaluasi Model

Evaluasi model dilakukan pada data uji 2025 untuk membandingkan performa model baseline dan model teroptimasi. Metrik utama yang digunakan untuk perbandingan adalah akurasi dan macro-F1, sedangkan analisis kesalahan diperkuat menggunakan *confusion matrix* ternormalisasi per baris. Sebagaimana model baseline pada Tabel 5, performa model teroptimasi turut dirinci melalui laporan klasifikasi lengkap untuk setiap kelas, sebagaimana disajikan pada Tabel 9.

Tabel 9. Hasil evaluasi model XGBoost teroptimasi pada data uji 2025

Metric	Precision	Recall	F1-Score	Support
0	0.97	0.74	0.84	225890.0
1	0.03	0.24	0.06	5416.0
2	0.53	0.79	0.64	38161.0
Accuracy	0.73	0.73	0.73	0.73
Macro Avg	0.51	0.59	0.51	269467.0
Weighted Avg	0.89	0.73	0.79	269467.0

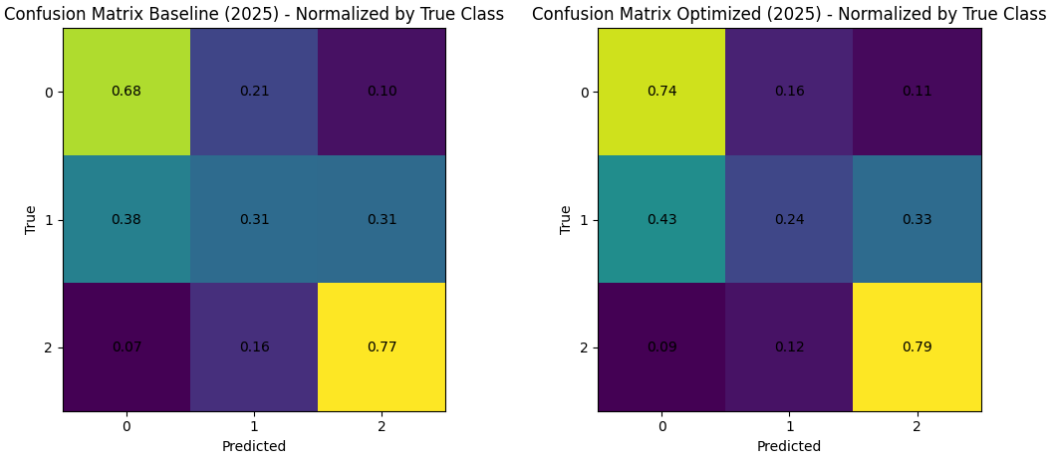
Jika dibandingkan dengan hasil baseline (lihat Tabel 5), model teroptimasi menunjukkan pola peningkatan yang tidak merata antar kelas. Pada kelas 0, recall meningkat dari 0,68 menjadi 0,74 dan F1 Score meningkat dari 0,80 menjadi 0,84, sementara precision relatif tidak berubah. Pada kelas 2, recall meningkat tidak terlalu signifikan dari 0,77 menjadi 0,79, namun precision sedikit menurun sehingga F1 Score-nya relatif stabil pada kisaran 0,64. Pada kelas 1 (pradiabetes) kelas yang menjadi fokus utama penanganan ketidakseimbangan pada penelitian ini, precision hampir tidak berubah dari 0,030 menjadi 0,031, sementara recall justru menurun

dari 0,31 menjadi 0,24, sehingga F1 Score-nya juga relatif stagnan dari 0,0548 menjadi 0,055. Kenaikan macro-F1 dari 0,499 menjadi 0,508 hampir seluruhnya disumbang oleh kenaikan F1 kelas 0, yang mengimbangi sedikit penurunan F1 kelas 2, sementara kontribusi kelas 1 terhadap kenaikan tersebut nyaris nihil. Temuan ini menegaskan bahwa peningkatan performa dari model optimasi Bayesian tidak berasal dari perbaikan kemampuan model mengenali kelas pradiabetes, melainkan didominasi oleh perbaikan pada kelas mayoritas.

Adapun ringkasan perbandingan kinerja model baseline dan teroptimasi ditampilkan pada Tabel 10. Model baseline mencapai akurasi 0,689 dan macro-F1 0,499, sedangkan model teroptimasi meningkat dengan akurasi sebesar 0,732 dan macro-F1 0,508. Peningkatan akurasi sebesar 4,3 poin ini sejalan dengan dominasi kelas 0 pada data uji yang sebanyak 83,8% dari total data, sehingga perbaikan recall kelas 0 memberikan pengaruh besar terhadap akurasi keseluruhan, sementara peningkatan macro-F1 sebesar 0,9 poin mencerminkan bahwa perbaikan tersebut tidak merata di seluruh kelas. Pola kesalahan prediksi divisualisasikan melalui *confusion matrix* ternormalisasi per baris model baseline dan teroptimasi pada Gambar 7.

Tabel 10. Perbandingan kinerja model baseline dan teroptimasi

Model	Akurasi	Macro-F1
Baseline	0.689	0.499
Teroptimasi	0.732	0.508



Gambar 7. Confusion matrix model baseline (Kiri) dan confusion matrix model teroptimasi

Visualisasi pada Gambar 7 memperkuat temuan pada Tabel 9, yaitu proporsi kasus pradiabetes yang terklasifikasi benar menurun dari 31% pada baseline menjadi 24% pada model teroptimasi, dengan proporsi kesalahan terklasifikasi sebagai tidak diabetes meningkat dari 38% menjadi 43%. Sebaliknya, proporsi klasifikasi benar pada kelas 0 meningkat dari 68% menjadi 74%, dan pada kelas 2 meningkat dari 77% menjadi 79%. Secara keseluruhan, hasil evaluasi model pada data uji 2025 menunjukkan bahwa model XGBoost teroptimasi memberikan peningkatan kinerja model dibandingkan baseline, namun peningkatan tersebut belum mencerminkan perbaikan pada tantangan utama penelitian ini, yaitu kemampuan untuk mengenali kelas pradiabetes secara akurat yang justru mengalami penurunan recall. Dengan demikian, pendekatan integrasi XGBoost dan Bayesian Optimization yang diterapkan dalam penelitian ini terbukti efektif meningkatkan performa model, tetapi belum sepenuhnya menjawab tantangan ketidakseimbangan kelas pada kategori pradiabetes, sehingga strategi tambahan yang lebih spesifik menyasar kelas minoritas diperlukan pada penelitian lanjutan.

KESIMPULAN

Penelitian ini menunjukkan bahwa integrasi XGBoost dan *Bayesian Optimization* pada dataset BRFSS 2023–2025, dengan skema *time-based split* dan prinsip *anti-leakage*, terbukti meningkatkan kinerja model prediksi diabetes dibandingkan dengan baseline. Pada data uji 2025, model teroptimasi mencapai akurasi 73,21% dan macro-F1 50,85%, lebih tinggi daripada baseline dengan akurasi 68,91% dan macro-F1 49,88%, sehingga menunjukkan efektivitas optimasi dalam memperbaiki performa generalisasi secara keseluruhan. Namun demikian, peningkatan tersebut didominasi oleh perbaikan pada kelas tidak diabetes, sementara recall pada kelas pradiabetes justru menurun dari 31% menjadi 24%, mengindikasikan bahwa optimasi hiperparameter belum berhasil menjawab tantangan ketidakseimbangan kelas minoritas dalam penelitian ini. Kelebihan penelitian ini terletak pada rancangan eksperimen yang reprodusibel dan terukur, termasuk hasil uji sensitivitas yang menunjukkan SMOTE tidak meningkatkan performa, bahkan cenderung menurunkannya. Keterbatasan utama penelitian ini adalah rendahnya kemampuan model dalam mengenali kelas pradiabetes secara konsisten, yang dimana tidak sepenuhnya teratasi dengan pembobotan kelas, sehingga penelitian lanjutan disarankan mengeksplorasi pendekatan *cost-sensitive* yang lebih spesifik menyasar kelas minoritas, strategi klasifikasi bertahap, serta evaluasi tambahan seperti kalibrasi probabilitas dan pengujian stabilitas lintas waktu untuk meningkatkan kinerja kelas minoritas tanpa mengorbankan generalisasi pada kelas mayoritas.

PENDANAAN

Penulis menyatakan bahwa penelitian ini tidak menerima dukungan pendanaan khusus dari lembaga pemerintah, swasta, maupun organisasi lainnya dalam pelaksanaan penelitian dan/atau publikasi artikel ini.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa penelitian ini dilakukan tanpa adanya hubungan komersial atau finansial yang dapat menimbulkan potensi konflik kepentingan.

KONTRIBUSI PENULIS

Penulis bertanggung jawab atas seluruh tahapan penelitian, termasuk konseptualisasi penelitian, penyusunan metodologi, pencarian dan seleksi literatur, analisis data, interpretasi hasil, penyusunan manuskrip, serta revisi akhir artikel. Penulis telah membaca dan menyetujui versi akhir manuskrip untuk dipublikasikan.

PERNYATAAN PENGGUNAAN KECERDASAN ARTIFISIAL GENERATIF

Penulis menyatakan bahwa kecerdasan artifisial generatif (Generative Artificial Intelligence/AI) tidak digunakan untuk menghasilkan data penelitian, melakukan analisis ilmiah, atau menggantikan kontribusi intelektual penulis dalam penyusunan manuskrip ini. Apabila digunakan dalam proses persiapan naskah, AI hanya dimanfaatkan secara terbatas untuk membantu penyempurnaan bahasa dan struktur penulisan. Penulis tetap bertanggung jawab penuh terhadap akurasi, integritas, serta isi akhir manuskrip.

REFERENSI

Alamsyah, N., Danestiara, V. R., Yoga, T. P., & Nursyanti, R. (2025). A Metaheuristic

Wrapper Approach to Feature Selection with Genetic Algorithm for Enhancing XGBoost Classification in Diabetes Prediction. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *10*(4). https://doi.org/10.22219/kinetik.v10i4.2366

Barman, B., Choudhury, H. K., & Jajodia, B. (2025). Interpretable Machine Learning for Diabetes Risk Prediction: A Large-Scale Analysis of Indian National Survey Data. *Discover Public Health*. https://doi.org/10.1186/s12982-025-01236-8

Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big Data Analytics in Healthcare. *BioMed Research International*, *2015*, 370194. https://doi.org/10.1155/2015/370194

Brahmandjati, A., Rahim, A. M. A., & Asharudin, F. (2024). Optimasi Prediksi Diabetes Dengan Algoritma XGBoost Dan Teknik Preprocessing Data. *Logic: Jurnal Ilmu Komputer Dan Pendidikan*, *3*(1), 116–125.

Cahyana, A., Susanto, E. R., & Parjito. (2025). Penerapan Algoritma XGBoost untuk Prediksi Diabetes: Analisis Confusion Matrix dan ROC Curve. *Fountain of Informatics Journal*, *10*(1). https://doi.org/10.21111/fij.v10i1.14311

Demirci, F. S., & Yıldırım, R. (2025). XAI-Powered Comparative Diabetes Prediction with XGBoost, SVM and Artificial Neural Networks. In *International Symposium on AI-Driven Engineering Systems* (pp. 111–114). https://doi.org/10.36287/setsci.22.30.001

Federation, I. D. (2025). *IDF Diabetes Atlas, 11th Edition* (11th ed.). International Diabetes Federation. https://diabetesatlas.org/

Fitriyadi, F., & Indriyati. (2025). Optimasi Integrasi Adaptif Metode Naive Bayes dan Information Gain untuk Prediksi Komplikasi Diabetes (Adaptive Integration Optimization of Naïve Bayes and Information Gain for Diabetes Complication Prediction). *Sistemasi: Jurnal Sistem Informasi*, *14*, 2990–3006. https://doi.org/10.32520/stmsi.v14i6.5666

Harefa, E. M., & Waruwu, L. P. (2025). Peningkatan Pengetahuan Penderita Diabetes Melitus Tentang Faktor Resiko Kejadian Diabetes Melitus Tipe II Melalui Pendidikan Kesehatan. *Jurnal Peduli Masyarakat*, *7*(3), 183–190. https://doi.org/10.37287/jpm.v7i3.6484

Khurshid, M. R., Manzoor, S., Sadiq, T., & Id, L. H. (2025). Unveiling Diabetes Onset: Optimized XGBoost with Bayesian Optimization for Enhanced Prediction. *PLoS ONE*, *1*, 1–29. https://doi.org/10.1371/journal.pone.0310218

Kliment, S., Information, F., & Technologies, C. (2018). Big Data Analytics in Medicine and Healthcare. *Journal of Integrative Bioinformatics*, 1–5. https://doi.org/10.1515/jib-2017-0030

Li, W., Id, Y. P., & Peng, K. (2024). Diabetes Prediction Model Based on GA-XGBoost and Stacking Ensemble Algorithm. *PLoS ONE*, 1–29. https://doi.org/10.1371/journal.pone.0311222

Mudjanarko, S. W., Pramono, R. B., & Mustikawati, D. E. (2023). Expert Opinion on Addressing the Gap in Injection Technique and Needle Reuse for People with Diabetes in Indonesia. *Diabetes, Metabolic Syndrome and Obesity*, 4101–4107. https://doi.org/10.2147/DMSO.S433254

Munshi, R. M. (2025). Optimising Hyperparameters with a Tree Structured Parzen Estimator to Improve Diabetes Prediction. *Scientific Reports*, 1–10. https://doi.org/10.1038/s41598-025-19295-x

Organization, W. H. (n.d.). *Diabetes*. World Health Organization.

Rizal, Azzahra, A. F., Bandaso, R. M. S., Harefa, J., & Jingga, K. (2025). Type 2 Diabetes Prediction Using Recall-Optimized XGBoost. *Procedia Computer Science*, *269*, 1309–1318. https://doi.org/10.1016/j.procs.2025.09.072

Shindy, T. N., Herajy, M., & Awad, W. A. (2025). Diabetes Detection Via Machine Learning: Tackling Challenges and Envisioning Future Innovations. *Alfarama Journal of Basic & Applied Sciences*, 412–430. https://doi.org/10.21608/ajbas.2025.391259.1259

Verma, V., Kumar, S., Verma, V. R., & Agrawal, A. (2025). An Intelligent Machine Learning Pipeline for Early Diabetes Prediction: CatBoost Ensemble with SMOTTEEN and Optuna Tuning. *International Journal of Computing and Artificial Intelligence*, *6*(2), 229–237. https://doi.org/10.33545/27076571.2025.v6.i2c.202